quinn emanuel trial lawyers | washington, dc
777 Sixth Street NW, 11th Floor, Washington, District of Columbia 20001-3706 | TEL (202) 538-8000 | FAX (202) 538-8100

WRITER'S DIRECT DIAL NO.
(202) 538-8166

WRITER'S INTERNET ADDRESS
mikelyle@quinnemanuel.com

February 13, 2015

Hon. Andrew J. Peck
United States Magistrate Judge, Southern
District of New York
Daniel Patrick Moynihan Courthouse
500 Pearl Street, Courtroom 20D
New York, New York 10007

Re: Rio Tinto v. Vale et al, Civil Action No. 14-cv-3042 (RMB) (S.D.N.Y.)

Dear Judge Peck:

Plaintiff Rio Tinto plc ("Plaintiff") and Defendant Vale S.A. ("Vale") write jointly to provide the Court with a revised proposed Predictive Coding Protocol. Pursuant to the Court's February 6, 2015 Order, the parties have continued to meet and confer with respect to certain suggested revisions to the proposed protocol and believe that we have resolved the Court's concerns with respect to various aspects of the protocol. In addition, the parties respective vendors have reviewed the proposed protocol and believe it is consistent with and can be applied to the parties respective predictive coding processes. While the proposed Predictive Coding Protocol requires the parties to exchange details about their respective predictive coding process, we also take this opportunity to provide the Court with a brief summary of those processes as provided by the parties' respective vendors.

Rio Tinto

As discussed at the parties February 6, 2015 hearing, Rio Tinto and its vendor, Precision Discovery, will be using Relativity Assisted Review ("RAR"). Using RAR, Rio Tinto will first create a Control Set by randomly sampling from the document universe. The legal subject matter expert will then review the control set for responsiveness. The Control Set is not used to train the set, it is only for validation. Following the control set review, Precision Discovery will note the Control Set's percentage of responsive documents. This number will serve as a benchmark throughout the project. It will also affect the size of the seed set. We will then

perform a seed set training review. The seed set may be created using random sampling, keyword searching, and/or conceptual ranking. After the first round of seed set review, Precision Discovery will use the coding from the seed set to categorize the document universe, including the control set. After categorization, the Control Set will have both computer coding and human coding. In comparing the human coding with computer coding, Precision Discovery will check for coding volatility, precision, recall, and F1 metrics to track progress and would expect to see an increase in precision, recall, and F1 from round to round. Volatility is also expected to decrease from round to round. These last two steps (the seed set training round and check for coding volatility, precision, recall and F1 metrics) will be repeated until Quinn Emanuel, in consultation with Precision Discovery, is satisfied with the metrics achieved. When the metrics of precision, recall and F1 are achieved, we will perform a last step of validation. For this final step, documents coded as non-responsive will be sampled. If the level of overturns within this sample is within the margin of error, the predictive coding project is complete. If it is above the margin of error, we will again repeat those same steps until the level of overturns is within the margin of error.

Vale

First, an initial Control Set will be created by drawing a statistically-valid random sample of documents from the review population. This will be used as the "gold standard" and should be coded by a member of the team who has a thorough understanding of the matter. Second, a Seed Set will be coded by a team of human reviewers. These documents can be selected at random or based on judgmental sampling. Third, Deloitte's Dynamic Review, which utilizes the LibLinear library as its document classification algorithm, will use the review team's coding to build a predictive model to categorize the rest of the documents in the Document Universe. Fourth, the predictive model will be used to assign a responsiveness score to the rest of the documents in the Document Universe. Fifth, the Control Set is used to determine the effectiveness of the model. Sixth, additional Training Sets are drawn from the Document Universe and reviewed. Documents in the Training Sets are generally drawn based on the responsiveness scores assigned by the model, but they can also be drawn at random or based on judgmental sampling. Seventh, in consultation with Deloitte, the legal team reviews the results of the model using precision and recall rates to make strategic decisions with respect to the unreviewed documents in the Document Universe. Depending on the results of this analysis, Steps 2 through 7 are repeated until precision and recall rates reach an appropriate level. Eighth, as a final validation, a randomly selected Validation Set will be pulled and reviewed to verify results are as predicted by the model.

****

The parties respectfully submit the attached proposed Predictive Coding Protocol for the Court's review.

Very truly yours,

/s/Michael Lyle
Michael Lyle